EXHIBIT O

OMNIBUS BROWN DECLARATION

A. COLIN CAMERON • PRAVIN K. TRIVEDI

MICROECONOMETRICS

METHODS AND APPLICATIONS

CAMBRIDGE

CAMBRIDGE

more information - www.cambridge.org/9780521848053

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521848053

First published in print format 2005

ISBN-13 978-0-511-12581-2 eBook (NetLibrary)
ISBN-10 0-511-12581-X eBook (NetLibrary)

ISBN-13 978-0-521-84805-3 hardback
ISBN-10 0-521-84805-9 hardback

7.6. Power and Size of Tests

The remaining sections of this chapter study two limitations in using the usual computer output to test hypotheses.

First, a test can have little ability to discriminate between the null and alternative hypotheses. Then the test has low power, meaning there is a low probability of rejecting the null hypothesis when it is false. Standard computer output does not calculate test power, but it can be evaluated using asymptotic methods (see this section) or finite-sample Monte Carlo methods (see Section 7.7). If a major contribution of an empirical paper is the rejection or nonrejection of a particular hypothesis, there is no reason for the paper not to additionally present the power of the test against some meaningful alternative hypothesis.

Second, the true size of the test may differ substantially from the nominal size of the test obtained from asymptotic theory. The rule of thumb that sample size $N > 30$ is sufficient for asymptotic theory to provide a good approximation for inference on a single variable does not extend to models with regressors. Poor approximation is most likely in the tails of the approximating distribution, but the tails are used to obtain critical values of tests at common significance levels such as 5%. In practice the critical value for a test statistic obtained from large-sample approximation is often smaller than the correct critical value based on the unknown true distribution. Small-sample refinements are attempts to get closer to the exact critical value. For linear regression under normality exact critical values can be obtained, using the t rather than z and the F rather than χ^2 distribution, but similar results are not exact for nonlinear regression. Instead, small-sample refinements may be obtained through Monte Carlo methods (see Section 7.7) or by use of the bootstrap (see Section 7.8 and Chapter 11).

With modern computers it is relatively easy to correct the size and investigate the power of tests used in an applied study. We present this neglected topic in some detail.

7.6.1. Test Size and Power

Hypothesis tests lead to either rejection or nonrejection of the null hypothesis. Correct decisions are made if H_0 is rejected when H_0 is false or if H_0 is not rejected when H_0 is true.

There are also two possible incorrect decisions: (1) rejecting H_0 when H_0 is true, called a **type I error**, and (2) nonrejection of H_0 when H_0 is false, called a **type II error**. Ideally the probabilities of both errors will be low, but in practice decreasing the probability of one type of error comes at the expense of increasing the probability of the other. The classical hypothesis testing solution is to fix the probability of a type I error at a particular level, usually 0.05, while leaving the probability of a type II error unspecified.

Define the **size of a test** or **significance level**

$$\begin{aligned} \alpha &= \Pr\left[\text{type I error}\right] \\ &= \Pr\left[\text{reject } H_0 | H_0 \text{ true}\right], \end{aligned} \tag{7.43}$$

with common choices of α being 0.01, 0.05, or 0.10. A hypothesis is rejected if the test statistic falls into a rejection region defined so that the test significance level equals the specified value of α. A closely related equivalent method computes the ***p*-value** of a test, the marginal significance level at which the null hypothesis is just rejected, and rejects H_0 if the p-value is less than the specified value of α. Both methods require only knowledge of the distribution of the test statistic under the null hypothesis, presented in Section 7.2 for the Wald test statistic.

Consideration should also be given to the probability of a type II error. The **power of a test** is defined to be

$$\begin{aligned}\text{Power} &= \Pr\left[\text{reject } H_0 | H_a \text{ true}\right] \\ &= 1 - \Pr\left[\text{accept } H_0 | H_a \text{ true}\right] \\ &= 1 - \Pr\left[\text{Type II error}\right].\end{aligned} \tag{7.44}$$

Ideally, test power is close to one since then the probability of a type II error is close to zero. Determining the power requires knowledge of the distribution of the test statistic under H_a.

Analysis of test power is typically ignored in empirical work, except that test procedures are usually chosen to be ones that are known theoretically to have power that, for given level α, is high relative to other alternative test statistics. Ideally, the **uniformly most powerful** (**UMP**) test is used. This is the test that has the greatest power, for given level α, for all alternative hypotheses. UMP tests do exist when testing a simple null hypothesis against a simple alternative hypothesis. Then the Neyman–Pearson lemma gives the result that the UMP test is a function of the likelihood ratio. For more general testing situations involving composite hypotheses there is usually no UMP test, and further restrictions are placed such as UMP one-sided tests. In practice, power considerations are left to theoretical econometricians who use theory and simulations applied to various testing procedures to suggest which testing procedures are the most powerful.

It is nonetheless possible to determine test power in any given application. In the following we detail how to compute the asymptotic power of the Wald test, which equals that of the LR and LM tests in the fully parametric case.

7.6.2. Local Alternative Hypotheses

Since power is the probability of rejecting H_0 when H_a is true, the computation of power requires obtaining the distribution of the test statistic under the alternative hypothesis. For a Wald chi-square test at significance level α the power equals $\Pr[W > \chi^2_\alpha(h)|H_a]$. Calculation of this probability requires specification of a particular alternative hypothesis, because $H_a : \mathbf{h}(\boldsymbol{\theta}) \neq \mathbf{0}$ is very broad.

The obvious choice is the **fixed alternative** $\mathbf{h}(\boldsymbol{\theta}) = \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is an $h \times 1$ finite vector of nonzero constants. The quantity $\boldsymbol{\delta}$ is sometimes referred to as the hypothesis error, and larger hypothesis errors lead to greater power. For a fixed alternative the Wald test statistic asymptotically has power one as it rejects the null hypothesis all the time. To see this note that if $\mathbf{h}(\boldsymbol{\theta}) = \boldsymbol{\delta}$ then the Wald test statistic becomes